DOCUMENT RESUME

ED 275 021 CS 505 403

AUTHOR Roberts, Charles V.
TITLE The Question of Validity and Application: How Do We Know How Well We Are Listening.
PUB DATE 86
NOTE 20p.; Paper presented at the Annual Meeting of the Eastern Communication Association (77th, Atlantic City, NJ, April 30-May 3, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Communication Research; Comparative Analysis; Higher Education; *Listening Comprehension Tests; Listening Habits; *Listening Skills; Long Term Memory; Short Term Memory; Test Reliability; *Test Validity
IDENTIFIERS Receiver Apprehension Test; *Watson Barker Listening Test

ABSTRACT

To test the validity of the Watson-Barker Listening Test (WBLT), a study was conducted based on the hypothesis that there is a curvilinear relationship between receiver apprehension and listening ability. Subjects, 120 undergraduate speech students, completed the Receiver Apprehension Test (RAT) and the WBLT. Analysis of the data revealed a significant correlation between RAT scores and both long-term memory and total listening ability, though not between RAT scores and short-term memory. These significant relationships were curvilinear in nature, supporting the claims of validity for the Watson-Barker instrument. However, conclusions drawn from the study emphasize that more data must be accumulated before the WBLT can be considered "externally valid." A five-page bibliography is appended. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275021

THE QUESTION OF VALIDITY AND APPLICATION: HOW DO WE KNOW HOW WELL WE ARE LISTENING

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Charles V. Roberts

McNeese State University

Lake Charles, Louisiana 70609

(318 - 437 - 5039)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles V. Roberts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the 1986 Convention of the Eastern Communication Association, Atlantic City, New Jersey

CS 505 403

ABSTRACT

THE QUESTION OF VALIDITY AND APPLICATION: HOW DO WE KNOW HOW WELL WE ARE LISTENING

The validity of the various listening tests have been established through inspection by listening theorists. This study sought additional support for these claims of validity. One hundred twenty students enrolled in basic speech courses were asked to complete the Receiver Apprehension Test (RAT) and take the Watson-Barker Listening Test: Form A. Statistical analysis of the data revealed a significant correlation between RAT scores and both Long Term Memory and Total Listening, but not between RAT scores and Short Term Memory. The significant relationships were curvilinear in nature, as expected, based on the relevant literature. It was concluded that the claims of validity for the Watson-Barker instrument are partially supported by this data.

The paper concludes with a general discussion of progress in listening research and pedagogical advances in the listening field.

[This material has been presented, by invitation, at SSCA in 1986.]

THE QUESTION OF VALIDITY AND APPLICATION: HOW DO WE KNOW HOW WELL WE ARE LISTENING

Listening is the most widely used human means of receiving information. Countless studies have verified this generalization (Rankin, 1926; Wilt, 1949; Breiter, 1957; and Duker, 1971). However, a concerted research interest in listening is relatively new. Duker (1964) mentions articles on listening that go back to the early 1900's, but few studies were actually undertaken prior to the late 1940's. After this date, interest increased dramatically. Pedegogical consideration of listening intensified during the fifties and early sixties, due in part, perhaps, to the research published by Nichols (1948) and Brown (1949). Scholars were aided immensely in these efforts by the availability of an instrument (Brown and Carlsen, 1955) that allowed for the diagnosing of listening comprehension skills. But almost as quickly as it grew, the interest in listening research declined.

In the mid-sixties a number of criticisms of listening tests, and indeed of the whole conceptualization of listening surfaced (See, for example, Becker, 1963; Petrie, 1964; Kelly, 1967). Perhaps for this reason, though the number of possible outlets for publishing research reports has increased dramatically since 1960, there has been less published research on listening in the last ten years then there was in the 1950's. This relative paucity of research is reflected in basic speech textbooks and, perhaps more critically, in the leading listening textbooks. Two of the most recent listening textbooks footnote as many studies done prior to 1960 as they do studies done after that date (Steil et al, 1983; Wolvin et al, 1982). Other scholarly works fare no better. In 1978, ERIC and SCA jointly issued Assessing Functional Communication (Larson et al, 1978) in which listening assessment was discussed. Only one of the references cited in the article was written within five years of the publication date of the article while nine were written prior to 1960.

Erway (1972), in attempting to explain why less research was being published, suggested that it was because it "has been difficult to measure valid changes in behavior because we have not yet decided what listening is" (p. 22). A valid and reliable test had not been agreed upon by a majority of the listening researchers. We continued to focus on measurement, to the detriment

of theory building. Perhaps it is with our listening research focus as Delia (1976) suggested it was with that of ethos and that "measurement procedures rather than theoretic explication " have been of more importance and thus have prevented real understanding of the conceptual area. It would seem more prudent to first discover what it is that we should be studying before deciding how we should measure it. Definitions are key building blocks for theory. Any definition of listening accepted by researchers would not only help shape their theories, but also would guide their investigations and suggest particular methodologies.

The focus on tests and not listening theory perhaps prompted Cronkhite (1974) to suggest that research be undertaken to investigate "variables that influence the audience's ability to reliably evaluate messages," and to "turn our existing speaker-oriented research upside down to discover implications for critical listening" (pp. 81-82). Sprague (1974) ,too, called for the translation of "speaker-oriented, control-oriented theories and research findings into receiver-centered, choice expanding implications"(p.83).

While the plea of such scholars for the creation of listening theories was persuasive, the successful translation of sender theories into receiver theories has not yet happened. It may be that it never will. Crucial to its success is the implied linkage between encoding and decoding. If we are to flip these theories over so that they address themselves to effective listening rather than effective speaking, should we not first ascertain if there is such a connection? So far such a connection has been suggested by many, accepted axiomatically by some, and substantiated by no published research that this writer has discovered. The fulcrum that would allow us this Atlas-like task remains elusive. We have yet to ascertain if listening and speaking mirror each other or "shadow" each other. Does one process reverse its opposite, repeat it, or are they totally different from each other? If they are reversible, then we can indeed "turn our... research upside down." But if the latter is the case, data derived theories of speaking "only" need be generalized "right-side up" to listening.

While there are indications that the decline in scholarly attention to listening is ending, it does not appear that our focus has wavered from the methodological question of how to best measure listening. The bulk of the current academic research effort seems to be concentrated on measurement rather than on the explication of listening theory. A new organization, the International Listening Association, was formed in the early 1980's to foster support for listening research and education. The business community has increased its emphasis on listening training. However, we still lack a conceptual delimination of the concept that would be

acceptable to a majority of the listening researchers within our discipline.

Differences in conceptualization and operationalization of variables abound in the literature. Such diversity is neither a surprise, nor a "curse." It allows for the emergence of the most robust theoretical explanation. Such definitional "battles" currently are being waged between various contending listening tests. The number of such listening assessment devices available to the modern communication researcher and/or teacher is increasing. Just as the Brown-Carlsen Listening Comprehension Test seemed to spur research activity during the ten years or so after its inception, various listening tests, especially the Kentucky Comprehensive Listening Test and the Watson-Barker Listening Test , seem to be prompting an increasing number of listening studies. But just as critics of the Brown-Carlsen Listening Comprehension Test questionned its validity, so too are there questions concerning the validity of these current measurement devices.

The Watson-Barker Listening Test was developed in 1982 in an attempt to create a standardized listening test that would be oriented primarily toward adults and mature college level audiences (Watson and Barker, 1984). The Kentucky Comprehensive Listening Test , also created in the early 1980's, seems similarly oriented. A number of reliability analyses have been conducted and acceptable levels of reliability established for both instruments. However, the only measure of validity undertaken for the Watson-Barker Listening Test was that of "face validity"(Watson and Barker, 1984, p.1). Given the diverse definitions of "listening" held by various listening experts, such support is not totally reassuring. Other efforts at establishing validity are being undertaken. Experiments are being conducted in an attempt to link test results of the Watson-Barker instrument with those of other listening tests such as the Kentucky Comprehensive Listening Test. While such experiments will help to establish the efficacy of comparing data of the various tests, they provide only a tautological validation of the instruments. If all tests are highly correlated and if any one test is valid, then the validity claims of all tests can be accepted. If no check of validity other than that of "face validity" is performed, all such claims should be held in abeyance until the concept of "listening" is agreed upon substantively by listening theorists.

Bostrom (1984) suggests that there are several ways of establishing validity and that the "usual definition (measures what it is 'supposed' to measure) does not exactly fit the kind of test that the KCLT represents" (p.2). He states that this is so because "each of the scales represents and [sic] actual instance of the performance of the skill in question"(p.2). Bostrom seems to be

avoiding the definitional battle by begging the question that while the definition of listening has not been agreed upon, the various subskills that his test measures have been accepted. If the "whole" has not been agreed upon, then the "parts" that make up the totality of that "whole "are no surer. While all agree either vigorously or ,at least tacitly, that certain subskills such as "retention" do belong within the province of listening, others are argued about vehemently. There is quite a bit of disagreement concerning which various subprocesses should be included within the conceptualization of listening. Is listening a combination of "hearing, understanding, and retaining" information, or should other subprocesses be included or some of these be excluded (Bostrom, 1984)?

The problems of establishing the validity of listening tests are monumental. Bostrom does discuss other methods of establishing validity. One procedure is to illustrate that the instrument in question measures a unique characteristic. He compares a wide variety of tests with his Kentucky Comprehensive Listening Test to illustrate its uniqueness. Brown (1985) reports similar tests that suggest that the Brown-Carlsen Listening Comprehension Test measures something different from reading comprehension, intelligence, and scholastic achievement. While this data is compelling evidence that these instruments measure unique constructs, it does not support the contention that they measure "listening ability." To say that something is not several other things is not the same as saying what it is.

Bostrom (1984) continues his quest for validity by illustrating that certain groups score differently than others on the test. Specifically he indicates that college students, army officers, and high school students have different performance levels. Bostrom suggests that "the KCLT does exactly what we might predict, showing different performance levels for each of these groups"(p.2). Knowing several members of each subject set, I suggest that none of the sets can boast of a uniform level of listening ability. Further, high school students out performed the other two groups in short-term listening and selective listening, while army officers scored better than both groups on lectures. I can find little theoretic or common sense backing for predictions in those directions. This is not to say that his instrument does not measure listening ability. Rather it is to suggest that he has not substantiated his case for the validity of his instrument using this criterion.

Regardless of the various conceptualizations of listening, it appears clear from the nature

of the instruments being used to measure "listening ability" that the one subprocess that is central to the measurement of listening is the "recall" or "recognition" of retained information. All tests share a common method. Subjects are asked to listen to a message, or set of stimuli, and then are asked to recall or recognize various parts of that message or set of stimuli either immediately after hearing the test passages or at some delayed time thereafter. While the nature of the test passage varies from instrument to instrument, this procedure seems invariant.

Another constant appears to be the effort on the part of the designers to hold "listening motivation" constant for all subjects. All of the major tests of listening ability are administered in such a manner so that all subjects are aware that their listening is to be tested. Kelly (1967) points out the problems of external validity using this procedure when he notes,

> We have a massive body of information about the listening behavior of subjects who knew they were going to be tested. . . but we have done almost nothing to find out about performances across the general range of situations from panic to boredom (p.464).

This is crucial to the external validity of listening tests when one considers that one of the most consistent findings in listening research has been that the recall of material is facilitated by increases in extrinsic motivational cues. Forewarning of a test has been found to be such a cue. Knowledge that a test will follow a listening experience has been labeled "anticipatory set." Anticipatory set creates the real possibility that a "ceiling effect" may be established. Procedures that are common in listening measurement severely limit the free functioning of any antecendent listening ability, as would be manifested in a "non-laboratory" situation. This phenomenon has been reported by many researchers (See, for example, Anastasis, 1961; and Kelly, 1962, 1965, 1967). Cronen and Mihevc (1972) discuss how subjects under "aware" conditions actively listen to messages so that they might answer questions concerning the material at a later time. The effect of forewarning is to raise the motivational forces naturally at work in the typical listener as high as his mental ability will allow and to disallow the differential functioning of other pertinent variables upon the comprehension and retention of material (Kelly, 1967). This may well be the reason that correlations between measures of mental ability and intelligence, and such listening tests as the Brown-Carlsen Listening Comprehension Test and the STEP have been so high (Keller, 1960; Petrie, 1961; Anderson and Baldauf, 1963).

Listening test designers should not be uninterested in studying the listening behavior of subjects under these conditions. Many classroom teachers hope that these conditions exist for them in their various courses. However, even a cursory inspection of the most ideal classroom will reveal that students are not motivated to listen, day in and day out, to the information presented them. Many students seem to be content to remember information only so long as it takes to place that information in their notes. In any case, conditions where testing is immanent are not likely to be found in most other situations.

Of particular interest then is the extent to which scores obtained in controlled conditions of standardized motivation reflect the listening ability of subjects when they venture outside the laboratory environment. Resolving this question of external validity is not an easy task, given the nature of the listening instruments extant today. While the Watson-Barker Listening Test does contain stimuli that are capable of being generated in a non-laboratory setting, the task of getting even one subject to respond to questions that would mirror the content of the test under conditions of "nonawareness of the intent to test" is too huge to seriously consider. The Kentucky Comprehensive Listening Test contains many items that would not be found outside of the laboratory (though the distracting stimuli contained in one part of the test could well be).

Although the dangers of the testing situation are obvious, research scholars are caught in a dilemma: if they warn subjects they are to be tested, the subjects are motivated (thus "standardizing" the test conditions and making inoperative many factors that would normally affect comprehension) and the test becomes artificial; yet, if subjects are not warned, reliability suffers and it cannot be considered a fair test. (Kelly, 1971, p.216)

At least one other method for severing the largely tautological Gordian Knot of validity claims was suggested by the efforts of Bostrom (1984). While uniqueness is one characteristic of validity, shared commonality, as evidenced by significant correlations with valid measures of a phenomenon is acceptable support of a contention of validity. There are tests of established validity that are conceptualized to measure certain aspects of the listening domain. One such instrument is the Receiver Apprehension Test (Wheeless, 1975). This instrument measures the self-reported anxiety of subjects that is associated with listening to stimuli generated in a variety of situations. It has been studied in terms of its relationship to other self-report measures (Beatty, 1981; Beatty and Payne, 1981) and its psychometric properties (Beatty, in press). Of particular note is the established correlation of RAT scores and physiological arousal (Roberts,

1980, 1984). This becomes even more important when the correlation between arousal and retention is entered into the equation. A number of researchers have established a link between retention and arousal (Kleinsmith and Kaplan, 1963; Crane et al, 1971; Roberts, 1980). The relationship between arousal and retention is posited to be curvilinear in nature, while the realtionship between physiological arousal and RAT scores is linear. Since listening ability is said to reflect short term retention and long term retention ability, in part, then there should be a correlation between RAT scores and scores on valid listening tests. This relationship would be curvilinear in nature. Too much or too little physiological arousal, as indicated by RAT scores, would result in poorer retention scores, as reflected by scores on a listening test. Optimum levels of arousal would result in higher retention scores.

In order to test the validity of the Watson-Barker Listening test, the following hypothesis was conceived:

> There is a curvilinear relationship between receiver apprehension, as measured by the RAT, and listening ability, as measured by the Watson-Barker Listening Test.

METHOD

SUBJECTS: Subjects were 127 volunteer undergraduate students, 42 males and 85 females, enrolled in beginning speech communication courses at a four-year university during the Spring semester of 1985. Data of seven of the subjects was subsequently discarded for several reasons. Three of the subjects were from other countries and their grasp of the English language prohibited an accurate test of their listening ability. Four other subjects did not complete one or both of the instruments utilized in this experiment.

PROCEDURE: At the beginning of the Spring semester, students in six sections of a basic speech communications class were asked to volunteer for an experiment. The purpose of the experiment was explained to them in detail and the procedures that would be followed were outlined. They were assured that the tests would have no impact on their grade, nor would their decision to participate or not participate affect their standing in the class. With only one exception, all students agreed to participate. The one non-volunteer was excused from the next class meeting.

At the next class meeting the subjects were asked to complete the Receiver Apprehension Test

(Wheeless, 1975). After collecting the RAT, subjects were asked to complete the Watson-Barker Listening Test, Form A (Watson and Barker, 1984). This test requires students to listen to a twenty minute audio tape and answer questions based on the information presented on the tape. There are five different types of listening tasks asked of the subjects. Each section of the test is comprised of ten questions. Three of the sections are said to test "short term memory skills" and the remaining two sections are purported to assess "long term memory skills"(Watson and Barker, 1984). The test tape begins with a short passage that allows the experimenter to insure that all subjects can hear the tape adequately. After adjusting the volume control of the tape player, the tape was played for the subjects, pausing only briefly to allow subjects to turn the pages of their test booklets when required. Although these pauses were not called for in the instructions provided with the test, they were deemed necessary because of the potential for distortion that the extraneous noise presented. The actual test time required varied slightly from class to class (the average time required for completing the Watson-Barker Listening Test was approximately 30 minutes). After the subjects had completed the test, their answer sheets were collected, they were asked to refrain from discussing the tests with others who might subsequently participate in the experiment, and were assured that their test answers would be evaluated, shared and explained to them at the next regular meeting of the class.

RESULTS

The completed tests were scored according to directions provided by the designers of the two instruments. As indicated above, four of the subjects failed to complete one or both of the tests and the tests of three othersubjects were discarded because it was evident that they did not understand English well enough to have their listening ability effectively measured by the Watson-Barker instrument. Pearson product-moment correlations were obtained for the scores of the remaining 120 subjects on the RAT and the Watson-Barker test measures of short term memory, long term memory, and total listening ability (short term memory plus long term memory). As suggested by the literature concerning the nature of the relationship between arousal, as tapped by the RAT instrument, and the retention dimension measured by listening tests, no significant relationships were established for total listening ability, short term listening, or long term listening (respectively the results were $r=.12$, $r=.13$, $r=.06$; $p>.05$).

While a certain level of arousal is necessary to perform cognitive tasks successfully, arousal levels beyond the optimum "readiness" level are dysfunctional (Cofer and Appley, 1964). As indicated above, previous research has shown that there is a significant linear correlation between RAT scores and physiological arousal. A direct relationship between memory and physiological arousal has been established as well. This relationship has been shown to be curvilinear in nature, in line with the "Activation Hypothesis" of Cofer and Appley. Since the Watson-Barker instrument does claim to measure retention, the relationship between it and the RAT most probably would not be linear in nature, but rather would be curvilinear in nature. The further the RAT scores are from the mean RAT score, the lower the Watson-Barker scores should be.

To test this proposed "inverted U-shaped" relationship, the 120 scores were arrayed on a scatter diagram and visually examined. This analysis strongly suggested that the relationship was not linear in nature. To statistically test this relationship the RAT scores of the 120 subjects were converted to absolute scores from the mean of the population (mean=40.89) and Pearson product-moment correlations were obtained for the adjusted RAT scores and the Watson-Barker scores of short term memory, long term memory, and total listening ability (Rosenthal and Rosnow, 1984, pp.222-224). Significant relationships were found to exist between the adjusted RAT scores and long term memory ($r=-.20$, $p<.03$) and between the adjusted RAT scores and total listening ability ($r=-.21$, $p<.02$), but not between the adjusted RAT scores and short term memory ($r=-.12$, $p<.18$). The power of the correlation test was .71 (Cohen, 1977).

DISCUSSION

The hypothesis was supported with regard to the relationships among the RAT scores and both long term memory and total listening ability, but not between short term memory and RAT scores. Previous researchers have suggested a strong link between arousal and long term retention, and a relatively weaker link between arousal and short term retention (Levonian, 1967; Roberts, 1980). These findings are in line with those results. Taken together with the previous literature on the arousal-retention relationship, this study provides evidence for the validity claims of the Watson-Barker Listening Test.

Establishing the validity of any new instrument is difficult. Given the relatively small portion of variance of listening scores that is accounted for by the RAT measure, definitive

conclusions concerning the validity of this new instrument must wait for additional data collection. Although the amount of variance accounted for is small, its magnitude is in line with Barker's (1984) conceptualization of listening which posits at least six different subprocesses as being involved with the listening process.

"Recall" is only one of these six processes and the only one to which the RAT has been empirically linked. It may well be that recall is of less importance than "attention," "hearing," "understanding," or any of the other possible subprocesses of listening, insofar as total listening scores are concerned.

However, this study does add weight to the claims of external validity for the Watson-Barker instrument. Further testing of the relationship between this listening test and measures of "attention," "understanding," etc., would help to increase confidence in this procedure. A more direct test of the relationship between listening scores on the Watson-Barker test and physiological arousal seems called for as well.

One additional note of caution is called for, based on the research project outlined above. While many claims of "face validity" have been made by the designers of listening tests, most of these tests seem, on the surface, to fail that test of validity because of the single medium nature of the test stimulus. Listeners generally do not "listen" with just their ears. Listening typically takes place while the listener is hearing and viewing the sender of the message. While attempting to assess the listener's ability to analyze the paralanguage message as well as the verbal message is indeed a useful pursuit, neglecting to measure the listener's ability to gain knowledge from the other aspects of nonverbal message transmission may render the total testing procedure useless in terms of applying the results to everyday encounters. Efforts are being undertaken to develop a listening test that more accurately measures the full range of decoding activities that the typical "listening" task involves. This new measurement procedure would include both the aural and the visual stimuli that are present in most communication situations. It is hoped that this new version of the Watson-Barker Listening Test will be found to be an even more valid and reliable measure of that nebulous concept we call listening.

That that research task should be undertaken before the definition battle outlined above is resolved is a moot point. The simple fact is that it will be done. The interest is there, and the need for such a tool is evident. Without it, we can not hope to develop effective methods for listening instruction. Given the rather sketchy evidence available, it is difficult to argue

with Erway's (1972) contention that gains from listening instruction are not maintained over time. Most listening research studies are "quick and dirty." Few longitudinal studies have been done. The generalizability of most studies is severely limited by the nature of the subject population drawn upon. By far the most prevalent educational level in listening research is the elementary school level. Fewer studies have been carried out at the secondary level, and only a handful have been completed using college-age subjects. This inverse relationship between the amount of studies and the age of subjects seems to mirror the relationship between age and potential for listening improvement that some researchers have alluded to in their articles (Evans, 1960; Everts, 1962; Lieb, 1965).

A close reading of most listening texts reveals that there is little reason to support the contention that we currently are effectively teaching listening. For such support we continue to have to fall back upon the subjective judgments of other teachers of listening. Erway (1972) has suggested, "the most impressive evidence comes not from research but from the prejudiced reports of students who have experienced instruction and from the observation of instructors" (p.23). This "evidence" must be considered especially suspect in light of the finding that people tend to think more highly of themselves as listeners than test scores indicate and that they are less able to discriminate between good and poor listening than they are between good and poor speaking (Stark, 1956). While it can be argued persuasively that we <u>should</u> teach listening at all educational levels, the only well documented listening finding is that listening <u>is not</u> being taught in most academic institutions.

Implementing longitudinal investigations that would document effective methods for teaching listening would help to reverse this tendency towards lip service. If scarcity does increase the value of a commodity, the results of such studies done in the classroom situation would prove very worthwhile. Prior to 1970, only <u>fifteen</u> empirical studies investigated pedagogical phenomena by first teaching teachers to behave in some particular way, then observing them to make sure they did behave in that way, and, finally, testing their students to note changes (Sprague, 1974). As noted previously, there are pronounced problems in generalizing laboratory research to the classroom and beyond. What is lost in terms of ability to control and limit experimental artifacts would be made up for in terms of the vigor and power of the generalizability of the resultant data.

Until such experiments are conducted, teachers interested in increasing listening skills can do no better than rely on the unsubstantiated platitudes that currently make up the bulk of our listening instruction. We will continue to tell our students to "Withhold evaluation of the message until the speaker is finished" (Barker, 1984, p. 55) and hope they don't ask us too many questions about the research that indicates that that is appropriate behavior. There is no research documentation that would support such imperatives. One even could argue that such a course of action is inefficient since it causes you to listen to unimportant as well as senseless drivel. Further, even if that inefficiency were shown to be necessary and/or useful, no pedagogical direction is available that would allow a teacher to help students carry out that directive. How does one "withhold evaluation" on the attitudinal level? Does the evaluation only matter if done on the "conscious" level? Does it matter if people do evaluate a speaker, if they still continue to listen to him?

The order in which research questions should be tackled is dictated, to a certain degree, by the urgency of situation. First we need to develop measures that are valid measures of listening, regardless of where and under what circumstances that activity takes place. Perhaps several instruments will be needed to cover all of the important contexts we wish to tap into. Then expediency necessitates that we undertake investigations to ascertain how we can best facilitate more effective listening. It may well be that our listening texts have more substance than alluded to above. If research reveals that there indeed are founts of knowledge and potent developers of skills already extant, more weight can be applied in the effort to wedge in listening instruction in our already crowded curricula. If none of our current teaching imperatives are supported, future research directions will be more clear and the weight of unsubstantiated dogma will no longer have to be borne by listening instructors. Which ever the case, we need to go forward.

> As long as we lack such research we shall be bound to myths and superstitions which are interesting subject matter for our methods courses, but which have little relevance for the real world (Sprague, 1974).

BIBLIOGRAPHY

Anastasis, A. Psychological Testing. New York: Academic Press, 1961.

Anderson, H. & R. Baldauf. A study of a measure of listening, Journal of Educational Research, 57 (1963), 197-200.

Barker, L. Communication. Englewood Cliffs, N.J.:Prentice Hall, Inc., 1984.

Beatty, M. Receiver apprehension as a function of cognitive backlog, Western Journal of Speech Communication, 45 (1981), 275-279.

Beatty, M. The effects of anticipated listening anxiety on receiver apprehension scores, Central States Speech Journal, in press.

Beatty, M. & S. Payne. Receiver apprehension and cognitive complexity, Western Journal of Speech Communication, 45 (1981), 363-369.

Becker, S. Relationships among the communicative skills, Central States Speech Journal, 14 (1963), 258-264.

Bostrom, R. The Kentucky Comprehensive Listening Test. Lexington: Kentucky Listening Research Center, 1983.

Bostrom, R. Research Update. Lexington: Kentucky Listening Research Center, 1984.

Breiter, L. R. Research in listening and its importance to literature. Master's Thesis, Brooklyn College, 1957.

Brown, J. The construction of a diagnostic test of listening comprehension, Journal of Experimental Education, 18 (1949), 139-146.

Brown, J. The Brown-Carlsen listening comprehension test, paper to be presented at the Speech Communication Association Convention, November, 1985.

Brown, J. and G. Carlsen. Brown-Carlsen Listening Comprehension Test, Manual of Directions, Yonkers-on-the-Hudson, New York: World Book Company, 1955.

Cofer, C. & M. Appley. Motivation: Theory and Research. New York: John Wiley and Sons, 1964.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.

Crane, L., R. Dieker, & C. Brown. Physiological responses to communication, in Listening: Readings, Vol. 2 (S. Duker, ed.), Metuchen, New Jersey: Scarecrow Press, 1971.

Cronen, V. & N. Michevc. Evaluation of deductive argument: A process analysis, Speech Monographs, 39 (1972), 124-131.

Cronkhite, G. Rhetoric, Communication and psychoepistemology: An academic exercise for the Sophistic devil, in Rhetoric: A Tradition in Transition (W. Fisher, ed.), East Lansing, Michigan: Michigan State University Press, 1974.

Delia, J. A constructivist analysis of the concept of credibility, Quarterly Journal of Speech, 62 (1976), 361-375.

Duker, S.(ed.) Listening Bibliography, New York: The Scarecrow Press, 1964.

Duker, S.(ed.) Listening Bibliography, second edition, Metuchen, New Jersey: The Scarecrow Press, 1968.

Duker, S.(ed.) Listening: Readings, Volume 2, Metuchen, New Jersey: The Scarecrow Press, 1971.

Erway, E. A. Listening: The second speaker, Speech Journal, 10 (1972), 22-27.

Evans, A. Listening related to speaking in the first grade, Master's Thesis, Atlanta University, 1960.

Everts, E. An investigation of the structure of children's oral language compared with silent reading, oral reading, and listening comprehension, unpublished Doctoral Dissertation, Indiana University, 1962.

Keller, P. Major findings in listening in the past ten years, Journal of Communication, 10 (1960), 29-38.

Kelly, C. Actual listening behavior of industrial supervisors, as related to "listening ability," general mental ability, selected personality factors and supervisory effectiveness, Unpublished Dissertation, Purdue University, 1962.

Kelly, C. An investigation of the construct validity of two commercially published listening tests, Speech Monographs, 32 (1965), 139-143.

Kelly, C. Listening: Complex of activities - and a unitary skill? Speech Monographs, 34 (1967), 455-465.

Kelly, C. Mental ability and personality factors in listening, Quarterly Journal of Speech, 49 (1963), 152-156.

Kleinsmith, L. & S. Kaplan. The interaction of arousal and recall interval in nonsense syllable paired-associate learning, Journal of Experimental Psychology, 65 (1963), 190-193.

Larson, C., P. Backlund, M. Redmond, & A. Barbour. Assessing Functional Communication, published jointly by Urbana, Illinois: ERIC Clearinghouse on Reading and Communication Skills and Falls Church, Virginia: Speech Communication Association, 1978.

Levonian, E. Retention of information in relation to arousal during continuously-presented material, American Educational Research Journal, 4 (1967), 103-116.

Lieb, B. The relationship between some aspects of communicative speaking and communicative listening in freshmen men and women. Master's Thesis, Pennsylvania State University, 1960.

Nichols, R. Factors in listening comprehension, Speech Monographs, 15 (1948), 154-163.

Petrie, C. An experimental evaluation of two methods for improving listening comprehension abilities, Unpublished dissertation, Purdue University, 1961.

Petrie, C. What we don't know about listening, Journal of Communication, 14 (1964), 248-252.

Rankin, P. T. The measurement of the ability to understand spoken language, Unpublished Dissertation, University of Michigan, 1926.

Roberts, C. A physiological validation of the receiver apprehension test, Communication Research Reports, 1 (1984), 126-129.

Roberts, C. The arousal-retention relationship in communication research, Paper presented at the Eastern Communication Association Convention, 1980.

Roberts, C. & T. Steinfatt. Source credibility and physiological arousal: An important variable in the credibility-information retention relationship, Southern Speech Communication Journal, 48 (1983), 340-355.

Rosenthal, R. & R. Rosnow. Essentials of Behavioral Research. New York: McGraw-Hill Book Company, 1984.

Sprague, J. Cognitive skills and teaching for mastery, in New Horizons for Teacher Education in Speech Communication (Newcombe, P. J. and R. R. Allen, eds.), Skokie, Illinois: National Textbook Company, 1974.

Stark, J. An investigation of the relationship of vocal and communicative aspects of speech competency with listening competency, unpublished Doctoral Dissertation, New York University, 1956.

Steil, L., L. Barker & K. Watson. Effective Listening: Key to Your Success, Reading, Massachusetts: Addison-Wesley Publishing Company, 1983.

Watson, K. & L. Barker. Watson-Barker Listening Test. New Orleans: Spectra, Inc., 1984.

Wheeless, L. An investigation of receiver apprehension and social context context dimensions of communication apprehension, Speech Teacher, 24 (1975),261-268.

Wilt, M. A study of teacher awareness of listening as a factor in elementary education, Unpublished Dissertation, Pennsylvania State University, 1949.

Wolvin, A. & C. G. Coakley. Listening, Dubuque, Iowa: William C. Brown Company Publishers, 1982.